PASSWORD STRENGTH ANALYZER USING LSTM AND CNN

Binoy Shiju, Anriya Jaison, Adinath Manoj, Arathy K L,Sanam E Anto, Ajith P J

Binoy Shiju,B.Tech Computer Science , Holy Grace Academy Of Engineering
Arathy K L,B.Tech Computer Science , Holy Grace Academy Of Engineering
Anriya Jaison,B.Tech Computer Science , Holy Grace Academy Of Engineering
Adhinadh Manoj,B.Tech Computer Science , Holy Grace Academy Of Engineering
Ajith P J, Project Guide, Holy Grace Academy Of Engineering
Sanam E Anto, Head of the Department(Computer Science),Holy Grace Academy Of Engineering

ABSTRACT

The Password Strength Analyzer using AI and Machine Learning is a web-based application designed to assess and enhance password security using a hybrid LSTM-CNN model.The system comprises a user-friendly frontend developed with HTML, CSS, and JavaScript, which allows users to input their passwords and receive instant feedback on their strength, score, and recommendations for improvement. The backend, built with Flask and TensorFlow, uses a pre-trained deep learning model that integrates Conv1D (CNN) for local pattern extraction and LSTM layers for capturing sequential dependencies, ensuring precise classification into Weak, Medium, or Strong categories. The model is trained on a large password dataset, tokenized at the character level, and uses padded sequences to maintain consistency. Additionally, the system performs real-time leak detection by querying the "Have I Been Pwned" API, checking if the password has appeared in known data breaches. The application also offers actionable recommendations, such as adding uppercase letters, digits, or special characters, to help users create stronger, more secure passwords. This tool provides an effective and interactive solution for promoting better password practices and strengthening online security.

Keyword :- *LSTM(Long Short Term Memory Machine), CNN(Convolution Neural Network), Flask, TensorFlow, Conv1D*

1.INTRODUCTION

In today's digital landscape, where data breaches and cyberattacks are increasingly common, the importance of strong password security cannot be overstated. Weak or compromised passwords remain one of the leading causes of unauthorized access and data theft. To address this issue, the Password Strength Analyzer using AI offers an intelligent solution for evaluating and improving password security. This project leverages artificial intelligence (AI) and deep learning to provide accurate and reliable password strength assessment, helping users create robust credentials to safeguard their sensitive information.The system utilizes a hybrid LSTM-CNN model to analyze password complexity effectively. The CNN (Conv1D) extracts local patterns from password sequences, while the LSTM layers capture long-term dependencies, allowing the model to classify passwords into three categories: Weak, Medium, and Strong. To further enhance security, the application integrates a real-time leak detection feature using the "Have I Been Pwned" API, which checks if the password has been involved in known data breaches.The frontend interface, built with HTML, CSS, and JavaScript, offers a user-friendly experience by allowing users to input their passwords and receive instant feedback on strength, score, and improvement suggestions. The backend, powered by Flask and TensorFlow, processes the input and returns the results efficiently. The model is trained on a diverse dataset of passwords, tokenized at the character level, and padded for consistency, ensuring accurate predictions.By providing actionable recommendations, such as incorporating uppercase letters, digits, and special characters, the Password Strength Analyzer empowers users to create stronger passwords. This project not only promotes better password practices but also enhances overall online security by offering a proactive defense against weak and compromised credentials.

2.LITERATURE SURVEY

Password-based authentication remains the most prevalent method despite known vulnerabilities stemming from users choosing weak, predictable, and often reused passwords based on dictionary words, personal information, or simple patterns. Research consistently highlights the tension between security requirements and user memorability, leading to easily guessable credentials. Early efforts to assess password strength relied on rule-based systems (checking length, character types like LUDS) and metrics like Shannon entropy, but studies indicate these often fail to capture true resistance against modern attacks like dictionary attacks, brute-force variations, and statistical guessing methods.

To address this, researchers developed more sophisticated approaches. Statistical methods like Markov models and Probabilistic Context-Free Grammars (PCFGs) were employed to model password structures and predict likely choices, outperforming simple dictionary lookups. Metrics evolved to include guessability (estimating cracking attempts) and linguistic distance (e.g., Levenshtein distance) to measure deviation from common words. Tools like zxcvbn emerged, incorporating pattern matching against leaked passwords, dictionaries, and common structures.

3.PROPOSED MODEL AND METHODOLOGY

The main goal of developing this model is to use machine learning methods to determine a password's strength.

3.1 Data Collection and Preprocessing

The dataset (passwords.csv) consists of labeled passwords categorized as Weak, Medium, or Strong. The data is preprocessed to remove duplicates, standardize text formats, and eliminate noise. Tokenization converts passwords into numerical sequences for training. Padding ensures uniform input length. Label encoding assigns numerical values to password categories. Special characters, numbers, and uppercase letters are analyzed to determine their impact on password strength. The cleaned dataset is split into training and testing sets for model development. Preprocessing is crucial for improving model accuracy, ensuring it learns meaningful patterns rather than random variations, leading to more precise password strength predictions. Fig 1:

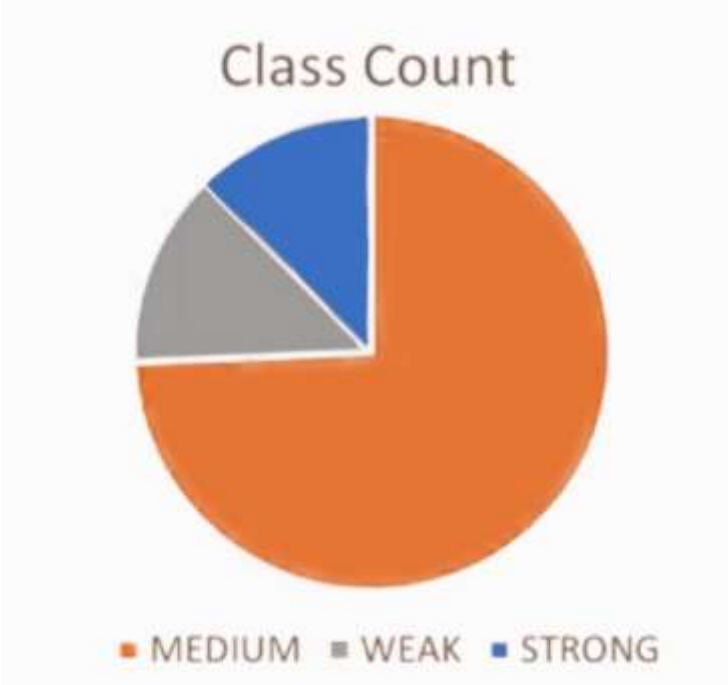

Fig 1: Class count for target attribute

3.2 Testing and training data

The dataset is divided into training (80%) and testing (20%) subsets to ensure generalization. The training data is fed into the machine learning model for learning password characteristics. The testing set evaluates performance, measuring accuracy, precision, recall, and F1-score. Overfitting is prevented through techniques like dropout layers and batch normalization. A validation set (split from training data) fine-tunes hyperparameters. The dataset includes a balanced mix of password strengths to avoid bias. The model is trained using an adaptive learning rate to optimize predictions. Proper data partitioning ensures the model effectively distinguishes weak, medium, and strong passwords.

4.MACHINE LEARNING CLASSIFIERS

Various classifiers can be used for password strength analysis, including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. CNN extracts spatial patterns in passwords but lacks sequential understanding. LSTM is better suited for password evaluation as it captures dependencies in character sequences. Both classifiers process tokenized passwords and classify them based on learned patterns. The study compares these models to determine which provides higher accuracy. CNN is faster but less effective for text-based sequences, while LSTM, designed for sequential data, outperforms in detecting strong and weak passwords based on previous character patterns within a password.

4.1 CNN Classifier

CNNs are primarily used in image processing but can be applied to text classification. The model consists of embedding layers, convolutional layers, pooling layers, and fully connected layers. Filters extract patterns like repeated characters or common weak password structures. The max-pooling layer reduces dimensionality while preserving important features. The final dense layer classifies passwords as Weak, Medium, or Strong. However, CNN struggles with sequential dependencies in passwords, limiting its effectiveness. While CNN offers computational efficiency, its inability to capture long-range dependencies in password structures makes it less suitable compared to models like LSTM for password strength classification.

Fig 2: Confusion Matrix for CNN

Table 1: Classification Report for CNN

	Precision	Recall	F1-Score	Support
0	0.98	0.98	0.98	14856
1	0.96	0.95	0.95	2459

2	0.99	0.99	0.99	2685
Accuracy			0.98	20000
Macro Avg.	0.98	0.97	0.97	20000
Weighted Avg.	0.98	0.98	0.98	20000

4.2 LSTM Classifier

LSTM networks are well-suited for analyzing passwords due to their ability to retain sequential dependencies. The model includes an embedding layer, LSTM layers, and dense layers for classification. The forget gate removes irrelevant information, while the input gate updates memory based on password patterns. The output gate determines final predictions. LSTM detects recurring patterns in weak passwords (e.g., "123456"), medium-strength patterns (e.g., "Passw0rd"), and strong passwords (e.g., "Ght@8Yz*#"). The model improves over time, learning password structures effectively. Its superior sequence learning capabilities make it ideal for classifying password strength based on learned dependencies from the dataset.

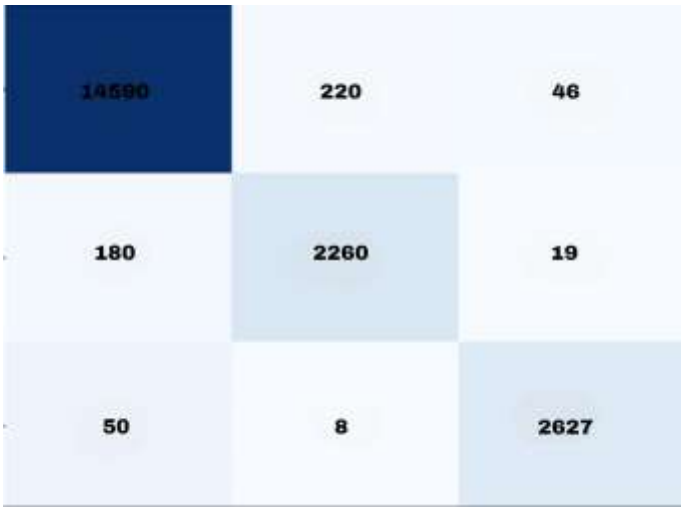

Fig 3: Confusion Matrix for LSTM

Table 2: Classification Report for LSTM

	Precision	Recall	F1-Score	Support
0	0.97	0.97	0.97	14856
1	0.94	0.92	0.93	2459
2	0.98	0.98	0.98	2685
Accuracy			0.97	20000
Macro Avg.	0.96	0.96	0.96	20000

Weighted Avg.	0.97	0.97	0.97	20000

5.RESULTS AND DISCUSSION

The model classifies passwords into three categories: Weak, Medium, and Strong based on learned features. Weak passwords often contain common words, repeated patterns, or short lengths. Medium-strength passwords have slight variations but remain predictable. Strong passwords use uppercase, lowercase, numbers, and symbols in random sequences. Model performance is evaluated using accuracy, precision, recall, and F1-score. LSTM achieves the best results due to its ability to capture sequential dependencies. The analysis reveals that weak passwords are still widely used, highlighting the need for awareness. Future improvements could involve training on a more diverse dataset for better generalization.

6.FUTURE RESEARCH DIRECTIONS

Future work could explore hybrid models combining LSTM with attention mechanisms or reinforcement learning to improve classification accuracy. Increasing dataset size and diversity can enhance performance. Adversarial training can be used to test the model's robustness against password guessing attacks. The integration of real-time password strength evaluation in browsers and password managers is another potential application. Future research could also focus on multi-factor authentication systems incorporating AI-driven password assessment. Additionally, exploring transformer-based models like BERT or GPT for password strength analysis could provide deeper insights into security vulnerabilities and improve classification accuracy further.

7.CONCLUSION

This study demonstrates the effectiveness of machine learning, particularly LSTM, in classifying password strength. Compared to traditional rule-based methods, deep learning models adapt better to complex password structures. The system accurately distinguishes weak, medium, and strong passwords, providing real-time feedback through a web-based interface. The research highlights the need for better password policies and AI-driven security solutions. While the model performs well, further improvements can enhance robustness. The study contributes to cybersecurity by promoting machine learning in password assessment, emphasizing the importance of strong password practices to mitigate security risks in digital environments.

More recently, machine learning (ML) has become central to password analysis. Supervised learning algorithms including Support Vector Machines (SVM), Decision Trees (DT), Random Forest (RF), Naïve Bayes (NB), Logistic Regression (LR), and XGBoost are frequently applied to classify passwords into strength categories (e.g., weak, medium, strong). Feature extraction is crucial, with techniques like TF-IDF vectorization converting passwords into numerical representations suitable for ML models.

Deep learning, particularly Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, offers advantages in handling the sequential nature of passwords, learning character dependencies, and predicting subsequent characters. Generative Adversarial Networks (GANs), like PassGAN, have been used to learn password distributions and generate realistic candidate passwords for guessing attacks or strength evaluation. Ensemble techniques such as Bagging and Stacking are also explored to combine multiple models, improving overall accuracy and robustness in password strength prediction. The importance of large, diverse, and sometimes language-specific datasets (often derived from real-world breaches like RockYou) is consistently emphasized for training and evaluating these models effectively, though challenges with dataset bias and quality persist. Password segmentation algorithms have also been proposed to analyze constituent parts of passwords. The overall trend moves towards data-driven, probabilistic, and ML/DL-based approaches for a more nuanced and accurate assessment of password security beyond simple compositional rules.

8.REFERENCES

[1]. Multi-Class Classification Prediction Model for Password Strength Based on Deep Learning -Seok Jun Kim1, Byung Mun Lee1 - march 07 2023 .

[2]. Analyzing Password Strength: A Combinatorial Entropy Approach - Naem Azam Chowdhury1 - 11 January 2024 .

[3]. Advancing user classification models: A comparative analysis of machine learning approaches to enhance faculty password policies at the University of Buraimi - Boumedyen Shannaq1, Oualid Ali2, Said Al Maqbali1, Afraa Al-Zeidi1 - 10 October 2024 .

[4]. A Conceptual Framework for Assessing Password Quality - Wanli Ma, John Campbell, Dat Tran, and Dale Kleeman - January 5 2007 .

[5]. Analysis Of Default Passwords In Routers Against Brute-Force Attack - Mohammed Farik, ABM Shawkat Ali- September 2015 .

[6]. A probabilistic Framework for Improved Password Strength Metrics - Javier Galbally, Iwen Coisel, Ignacio Sanchez - 30 march 2016 .

[7]. Analyzing Password Strength - Martin M.A. Devillers - July 2010

[8]. Exploratory Data Analysis on Username-Password Dataset - Vanita Jain1, Rishab Bansal2 and Mahima Swami - DOI: 10.5281/zenodo.5169881 - may 10 2021 .

[9]. Strength Analysis of Real-Life Passwords Using Markov Models - Viktor Taneski , Marko Kompara, Marjan Hericko and Boštjan Brumen - 30 september 2021 .

[10]. Deep Learning for Password Guessing and Password Strength Evaluation, A Survey - Tao Zhang134, Zelei Cheng2, Yi Qin, Qiang Li, Lin Shi - December 2020 .

[11]. Password strength verification based on machine learning algorithms and LSTM recurrent neural networks - Vladimir V. Belikov, Ivan A. Prokuronov - 2 may 2023 .

[12]. Machine-Learning-Based Password-Strength-Estimation Approach for Passwords of Lithuanian Context - Ema Darbutaite, Pavel Stefanovic and Simona Ramanauskaite - 30 june 2023 .

[13]. Measuring Password Guessability for an Entire University - Michelle L. Mazurek, Saranga Komanduri, Timothy Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Patrick Gage Kelley , Richard Shay, and Blase Ur - 13 november 2013 .

[14]. Password Strength Analyzer Using Segmentation Algorithms - Sivapriya K , Deepthi L.R -November 23 2020 .

[15]. Password Strength Meters: Implementations and Effectiveness - Dalton Gusaas - December 5 2015.

[16]. Real Time Password Strength Analysis on a Web Application Using Multiple Machine Learning Approaches - Umar Farooq - December 2020 .

[17]. An Investigation of Machine Learning for Password Evaluation - Margaret Nicole Todd - November 2016 .

[18]. Proactive Password Strength Analyzer Using Filters and Machine Learning Techniques - Suganya G , Karpgavalli S, Christina V - October 2010 .

[19]. An Online neural network based password prediction, generation and storage scheme. - Mbaka, Winnie Bahati - September 2021 .

[20]. Password Strength Analyzer - Bhavani Gorle - November 2022 .

[21]. Password Strength Analysis and its Classification by Applying Machine Learning Based Techniques - Sakya Sarkar , Mauparna Nandan - august 2022 .

[22]. Enhancing Multi-Class Password Strength Prediction Through Machine Learning and Ensemble Techniques - Enas F. Aziz , Mohammed Rashad Baker - 15 october 2024 .

[23]. Predicting The Strength of Password Using ML - RAJESH AREPALLI, G. SUPRIYA, SK. SHEEMA, A.L.S. DEVI, D. GOVARDHAN, G. SUBBARAO - July 2024 .

[24]. Machine Learning Based Password Strength Analysis - Sony Kuriakose, G Krishna Teja, Sravan Duggi, A Harshel Srivatsava, Venkat Jonnalagadda - July 2022 .

[25]. Enhancing Password Security and Memorability Using Machine Enhancing Password Security and Memorability Using Machine Learning and Linguistic Patterns Learning and Linguistic Patterns - jared wise - december 2024.